Harmonic Hidden Markov Models for the Study of EEG Signals

Bruno Torrésani, Emilie Villaron

HARMONIC HIDDEN MARKOV MODELS FOR THE STUDY OF EEG SIGNALS

Bruno Torrésani, Emilie Villaron

Université de Provence, LATP, CMI,
39 rue Joliot-Curie, 13453 Marseille Cedex 13, France
email: Bruno.Torresani@cmi.univ-mrs.fr Emilie.Villaron@cmi.univ-mrs.fr
web: www.latp.univ-mrs.fr

ABSTRACT

A new approach for modelling multichannel signals via hidden states models in the time-frequency space is described. Multichannel signals are expanded using a local cosine basis, and the (time-frequency labelled) coefficients are modelled as multivariate random variables, whose distribution is governed by a (hidden) Markov chain. Several models are described, together with maximum likelihood estimation algorithms.

The model is applied to electroencephalogram data, and it is shown that variance-covariance matrices labelled by sensor and frequency indices can yield relevant informations on the analyzed signals. This is examplified by a case study on the characterization of alpha waves desynchronization in the context of multiple sclerosis disease.

1. INTRODUCTION

Many signal classes exhibit specific features that are both time and frequency localized. This motivates the use of time-frequency signal representations, such as short time Fourier or Gabor transforms (see [2, 4, **?**] and references therein), or orthonormal bases such as MDCT or Wilson bases. Such elementary waveform systems have more recently been exploited in sparse regression algorithms (matching pursuit and variants, basis pursuit,...) in various contexts. However, real signals often exhibit significant correlations between time-frequency coefficients, which are often poorly accounted for by classical sparse expansion methods. In the case of multichannel signals, such as EEG, MEG,... signals, correlations also exist between channels: information may be somewhat localized in the channel domain, but close channels (sensors) are often significantly correlated.

Sparse regression approaches are often not very good at describing time-correlated or frequency-correlated data, contrary to channel correlations for which dedicated techniques have been developed (see for example [5, 6, 10] and references). Motivated by applications to EEG/MEG signal analysis, we discuss here a stochastic model implementing simultaneously channel and frequency correlated signals, following the lines of [**?**, 1], with explicit modelling of non-stationarity. Indeed, standard stochastic signal models often rely on (at least implicit) stationarity assumptions. However localized phenomena in EEG signals such as alpha or gamma oscillations cannot be accounted for with such assumptions.

The approach we propose models the distribution of time-frequency coefficients (here MDCT coefficients) in terms of latent variables. The latter control the covariance matrices of Gaussian vectors of fixed-time vectors of multi-channel, multi-frequency, MDCT coefficients. In the framework of application to EEG signals, the latent variables describe some hidden mental state of the subject. This model may be seen as a way to introduce non-stationarity in approaches such as that of [**?**, 1], where covariance matrices are expanded as linear combinations of tensor products of frequency and channel matrices. Inference for the proposed model can be done using fairly classical algorithms, which yield estimates for the model parameters, together with maximum likelihood estimates for the sequences of latent variables.

We illustrate the model and algorithms with a case study of EEG signal analysis, namely the detection of alpha waves in rest EEG for multiple sclerosis patients and control subjects. We show that the latent variable estimation is precise enough to confirm a biological hypothesis, namely the existence of inter-hemispherical alpha wave desynchronization on multiple sclerosis subjects, which may be a potential target for diagnosis.

2. HARMONIC HIDDEN MARKOV MODEL

We describe in this section a stochastic model for vector-valued (i.e. multichannel) signals, based on a local Fourier basis expansion with random coefficients, and discuss the corresponding inference problem. Throughout this paper, we denote by $\mathscr{B} = \{u_{tf}, (t,f) \in \Lambda\}$ a frame of time-frequency atoms, to be used to expand signals. We shall limit ourselves here to finite-dimensional local Fourier bases (also called MDCT bases, see e.g. [**?**] for a summary), of the form

$$u_{tf}[n] = \sqrt{\frac{2}{L}}\, w[n-tL] \cos\left(\pi(f+1/2)\frac{n-tL}{L}\right) , \qquad (1)$$

$t = 0,\dots N_t - 1$ (resp. $f = 0,\dots N_f - 1$) denoting the time (resp. frequency) index. Here w is a window function, subject to compatibility relations that ensure the family to be an orthonormal basis of the signal space $\mathscr{H} = \mathbb{R}^{N_t N_f}$. Notice that the model is straightforwardly extended to Gabor frames (in which case the estimation algorithms have to be modified though).

2.1 The model

Consider a time-frequency basis $\mathscr{B} = \{u_{tf}\}$, for example a MDCT basis, and N_c signals (representing N_c sensors's observations) of the form

$$\underline{x}^c = \sum_{t,f} a^c_{tf} u_{tf} \qquad (2)$$

The (vector) synthesis coefficients $\underline{a}^c_{tf}$ are modeled using hidden Markov models (see [8] for a review), as dependent Gaussian random vectors, whose distribution is governed by a time dependent latent state $X_t \in \{1,\dots N_s\}$. More precisely, set $\underline{A}_t = \{(a^c_{tf}, f = 0,\dots N_f - 1), c = 1,\dots N_c\}$. Then we assume that

- Conditional to $X = \{x_t, t = 0,\dots N_t - 1\}$ the vectors $\underline{A}_s$ are mutually independent zero mean Gaussian vectors, with density

$$\psi_{\underline{A}_t}(\underline{a}|X_t = k) = \mathscr{N}(\underline{a}; 0, \Sigma_k)$$

- The latent states (which depend on the time index only, and are common to all channels) form a Markov chain, characterized by its transition matrix π.

In addition, for avoiding working with large dimensional problems, we shall be particularly concerned with models in which the covariance matrix takes the form of a Kronecker product

$$\Sigma = \Sigma^{(c)} \otimes \Sigma^{(f)} ,$$

$\Sigma^{(c)}$ (resp. $\Sigma^{(f)}$) representing channels (resp. frequencies) covariances. More precisely, we shall assume that

$$\mathbb{E}\left\{a_f^c.a_{f'}^{c'}\right\} = \Sigma_{cc'}^{(c)}\Sigma_{ff'}^{(f)} ,$$

where the time index has been dropped for the sake of simplicity. This is very much in the spirit of [**?**, 1]. Notice that such a Kronecker product form is not unique, i.e. given Σ_k, $\Sigma_k^{(c)}$ and $\Sigma_k^{(f)}$ are defined up to a multiplicative factor. This fact has to be taken care of in the estimation procedure.

2.2 Inference

2.2.1 EM algorithm

Since we are using here an orthonormal basis, the inference problem can be solved using classical tools, after numerical computation of the coefficients $a_{tf}^c = \langle \underline{x}^c, u_{tf}\rangle$ using the freely available matlab toolbox LTFAT [9].

EM is an efficient iterative procedure to compute the Maximum Likelihood (ML) estimate in the presence of hidden data. In ML estimation, we wish to estimate the model parameters for which the observed data are the most likely. For the sake of completeness we describe here the main points of the algorithm. Each iteration EM involves two processes: The E-step (Expectation), and the M-step (Maximization). In the first one, the log-likelihood of the observed data is estimated given the current estimate of the model parameters, using the forward and backward variables defined here under. In the M-step, the likelihood function is maximized leading to the so-called Baum-Welch re-estimation formulas. Convergence is ensured since the algorithm is guaranteed to increase the likelihood at each iteration.

Expectation: For $t \in 0,..,N_t-1$, we define the following quantities : the forward variables $\alpha_t = (\alpha_t^k)_{k=1,..,N_s}$ are the normalized distribution of the latent state X_t conditional to the observed coefficients $(\underline{A}_r)_{r=0,...t}$:

$$\alpha_t^k = \mathbb{P}(X_t = k/(\underline{A}_r)_{r=0,...t}) \times L_t$$

where L_t is the likelihood of the observations until time t. The backward variable β_t^k is the likelihood of the observations $(\underline{A}_{t+1}, \underline{A}_{t+2}, ..., \underline{A}_{N_t-1})$ conditional to $X_t = k$ for $k = 1,\dots N_s$ and $t = 0,\dots N_t-1$.

The forward variables may be obtained recursively for $t = 1,\dots N_t-1$ while the backward ones may be achieved by recursion too, operating on decreasing indices $t = N_t-1$ down to 0. Thanks to these recursive equations, the complexity of the computation is linear in t.

Given the forward and backward variables, the distribution of the transition (X_t, X_{t+1}) for $t = 0,\dots N_t-2$ conditional to the observations up to final time N_t-1 reads

$$\mathbb{P}(X_t = k, X_{t+1} = l/(\underline{A}_t)_{t=0:N_t-1}) = \frac{1}{L_{N_t-1}}\alpha_t^k \pi_{k,l}\psi(\underline{A}_{t+1})\beta_{t+1}^l \quad (3)$$

where

$$L_{N_t-1} = \sum_{k=1}^{N_s} \alpha_t^k \beta_t^k$$

for every time t under consideration.

As a consequence, the distribution of any hidden state X_t, $0 \le t \le N_t-2$ conditional to the observations up to final time N_t-1 satisfies

$$\mathbb{P}(X_t = k/(\underline{A}_r)_{r=0:N_t-1}) = \frac{1}{L_{N_t-1}}\alpha_t^k\beta_t^k \quad (4)$$

In order to avoid underflow potentially caused by very small values of the probabilities, we used normalized versions of α and β defined by Rabiner in [8], to which we refer for more details.

Maximization: the maximum likelihood estimates for the model parameters read:

$$\widehat{\nu_k} = \frac{\alpha_0^k\beta_0^k}{L_{N_t-1}} \quad (5)$$

$$\widehat{\pi_{k,l}} = \pi_{k,l}\frac{\sum_{t=0}^{N_t-2}\alpha_t^k \times \psi_l(\underline{A}_{t+1}) \times \beta_{t+1}^l}{\sum_{t=0}^{N_t-2}\alpha_t^k\beta_t^k} \quad (6)$$

$$\widehat{\Sigma_k} = \frac{\sum_{t=0}^{N_t-1}\alpha_t^k\beta_t^k\underline{A}_t\ {}^t\underline{A}_t}{\sum_{t=0}^{N_t-1}\alpha_t^k\beta_t^k} \quad (7)$$

2.2.2 Estimation for tensor product covariance matrices.

In order to control the curse of dimensionality when all sensors are kept (from 32 to 64 sensors are usually used in EEG), we also consider models in which covariance matrices take the form of Kröneker products of frequency and channel matrices.

$$\Sigma_k = \Sigma_k^{(c)} \otimes \Sigma_k^{(f)} , \quad k = 1,..,N_s$$

This approach can be seen as a refinement of the general decomposition of the covariance into a sum of Kronecker products introduced by Bijma and De Munck in [**?**, 1]. The above approaches can be modified in order to account for that particular structure.

Expectation: the E-step remains unchanged but the computation of the density simplifies, owing to the fact that

$$\begin{aligned}\Sigma_k^{-1} &= (\Sigma_k^{(c)})^{-1} \otimes (\Sigma_k^{(f)})^{-1} \\ \det(\Sigma_k) &= \det(\Sigma_k^{(c)})^{N_f} \times \det(\Sigma_k^{(f)})^{N_C}\end{aligned}$$

Maximization: the re-estimation of covariance matrices in M-step is achieved using an alternating minimization procedure : consider $\underline{A}_t^c = (\underline{a}_{t,f}^c, f = 0,\dots N_f-1)$ and $\underline{A}_t^f = (\underline{a}_{t,f}^c, c = 1,\dots N_c)$ and iterate the following steps

- *Estimation of* $\Sigma^{(c)}$ *given* $\Sigma_k^{(f)}$*:* let $\Sigma_k^{(f)} = ({}^tB_kB_k)^{-1}$ denote a Cholesky decomposition of $\Sigma_k^{(f)-1}$, define

  $$M_t^k(c,c') = \langle B_k\underline{A}_t^c\ , B_k\underline{A}_t^{c'}\rangle$$

  and set

  $$\widehat{\Sigma_k^{(c)}} = \frac{1}{N_f}\frac{\sum_{t=0}^{N_t-1}\mathbb{P}(X_t=k)M_t^k}{\sum_{t=0}^{N_t-1}\mathbb{P}(X_t=k)} \quad (8)$$

- *Normalization*: set

  $$\widehat{\Sigma_k^{(c)}} = \widehat{\Sigma_k^{(c)}}/\|\widehat{\Sigma_k^{(c)}}\|_F , \quad (9)$$

  $\|\cdot\|_F$ denoting the Frobenius norm.

- *Estimation of* $\Sigma^{(f)}$ *given* $\widehat{\Sigma_k^{(c)}}$*:* Using the Cholesky decomposition $\widehat{\Sigma_k^{(c)}} = ({}^tD_kD_k)^{-1}$, define

  $$P_t^k(f,f') = \langle D_k\underline{A}_t^f, D_k\underline{A}_t^{f'}\rangle$$

  and set

  $$\widehat{\Sigma_k^{(f)}} = \frac{1}{N_c}\frac{\sum_{t=0}^{N_t-1}\mathbb{P}(X_t=k)P_t^k}{\sum_{t=0}^{N_t-1}\mathbb{P}(X_t=k)} \quad (10)$$

These estimators are obtained by alternate optimization of the log-likelihood with respect to $\Sigma_k^{(c)}$ and $\Sigma_k^{(f)}$ respectively. The normalization step described here above enables us to raise an indetermination since $\Sigma_k^{(f)}$ and $\Sigma_k^{(f)}$ are defined up to a constant. Strictly speaking, this corresponds to a GEM algorithm for which the convergence to local minimum is still ensured by the fact that each step increases the likelihood of the observations with respect to the model.

Remark 1 As an alternative, the full covariance matrix Σ can also be estimated classically, the channel and frequency covariances being estimated afterwards by factorization, i.e. by minimizing a mean square error $\left\|\Sigma_k - \Sigma_k^{(c)} \otimes \Sigma_k^{(f)}\right\|_F^2$ under the constraint $\left\|\Sigma_k^{(c)}\right\|_F = 1$.

2.2.3 Estimation of the hidden states sequence

Giving the estimated MAP parameters, we use the Viterbi dynamic programming algorithm to find the most likely sequence of hidden states

$$(X_0^{MAP},...,X_{N_t-1}^{MAP}) = \underset{l=(l_0,..l_{N_t-1})}{\arg\max}\ \mathbb{P}(X = l|\underline{A}_t, t = 0,..,N_t-1) \quad (11)$$

We used the algorithm detailed in [8]. This procedure is also linear in t but requires numerical evaluation of probability densities, i.e. inversion of covariance matrices. Therefore, when N_f and N_c are large, using tensor product decompositions yields more stability in the numerical evaluation.

2.3 Numerical simulations

To validate the algorithms, we simulated signals ($N_c = 8$ channels of length 100000 samples each, corresponding to 40 sec of signal) following the above model.
For the sake of simplicity, and because of the nature of the real EEG signals we focus on in section 3, we limite ourselves to $N_s = 2$ hidden states. MDCT transform was tuned so as to generate $N_f = 10$ frequency bins for each channel.
The frequency covariance matrices were set to the estimated values for the EEG case study presented below (see Fig. 4 in section 3.2) and artificial channel covariance matrices were generated so as to exhibit topographically distinct sources for the two states (see Fig. 1).

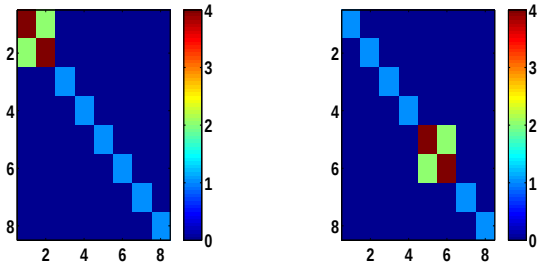

Figure 1: Channel covariance matrices for state 'non-alpha' (left) and 'alpha' (right)

A reference transition probability $p_{11} = \mathbb{P}(X_t = 1|X_{t-1} = 1) = 0.85$ was chosen (corresponding to a typical value for alpha waves duration, i.e. 700ms), simulations were run for various values of $p_{00} = \mathbb{P}(X_t = 0|X_{t-1} = 0)$ in the interval $[0.75; 0.95]$ (100 runs per value of p_{00}), and relative errors $\frac{\|\hat{\Sigma}-\Sigma\|_F}{\|\Sigma\|_F}$ were computed.
The results, displayed in Fig. 2, clearly show that the Kronecker product based estimate outperforms the classical one in terms of accuracy (for comparable running time).
As expected, when p_{00} grows, the average length of zero-state regions increases, and the accuracy of the estimate of Σ_0 (resp. Σ_1) grows (resp. decays).
The hidden state identification error rates are very comparable for the two methods and always remain at a very low level (less than 0.5%).

3. APPLICATION TO EEG SIGNALS

The above model, inspired by [7], was motivated by applications to the modeling of EEG signals, and we describe now a case study,

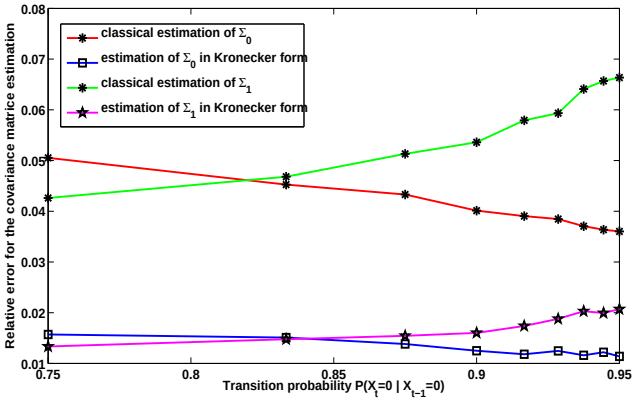

Figure 2: Evolution of the estimation relative errors as a function of the transition probability p_{11}

devoted to detection of alpha waves in rest EEG signal[1].

Alpha waves are short duration time-localized oscillations (with frequency around 10 Hz) which appear in EEG signals in some specific situations, and can naturally be accounted for by time-frequency representations. They are also topographically localized in specific sensors situated in posterior regions of head. The alpha wave occurrence may be considered a non-stationary effect, i.e. a departure from a stationary background signal. This therefore motivates the use of hidden Markov models as described above.

3.1 Problem statement and data

We aim at automatic detection of alpha waves, i.e. segmentation of the signal into time epochs where alpha waves are present/absent. Such a task could be achieved by band pass filtering of signals, followed by appropriate thresholding, or time-frequency transform thresholding. In both cases this would however raise the question of the determination of the threshold. The HHMM described above provides a more systematic approach.

The main goal of the current study is to use the model in an unsupervised manner, in order to test its ability to blindly detect alpha waves. The case study below provides a positive answer. In other situations though (i.e. with more complex signals), supervised training may be necessary in order to obtain accurate detection.

For the problem under consideraton, we use the model with two instances for the hidden states: alpha ($k = 1$)/non-alpha ($k = 0$) whose parameters are

- the initial hidden state distribution ν where
$$\nu_k = \mathbb{P}(X_0 = k)$$
- the stochastic transition matrix of the Markov Chain Π where
$$\Pi_{k,l} = \mathbb{P}(X_{t+1} = l|X_t = k)$$
- the covariance matrices $\Sigma_0^{(f)}, \Sigma_1^{(f)}, \Sigma_0^{(c)}, \Sigma_1^{(c)}$.

Once the estimation of these parameters is achieved using the EM algorithm described here above, we determine the most likely hidden states sequence. This sequence is the main ingredient in the case study below since it detects the bursts in the time-frequency images which corresponds to alpha waves. Indeed, it is clearly seen in Fig. 3 that values equal to one correspond to regions in the time-frequency domain with significantly larger MDCT coefficients in the alpha band.

[1] Let us stress however that a similar approach can be adapted to much more general situations, in which signals exhibit sufficient time-frequency localized components.

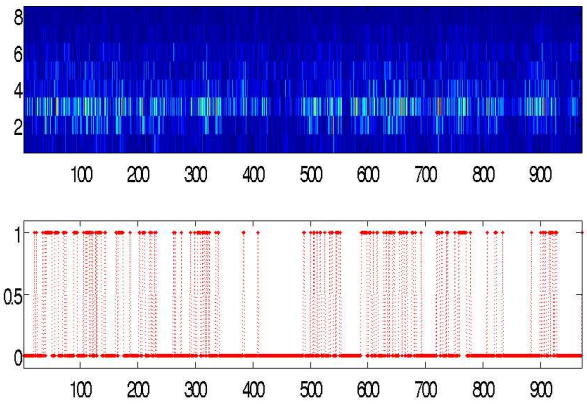

Figure 3: MDCT coefficients for a fixed sensor (top) and hidden states sequence estimated (from all sensors) using Viterbi algorithm (bottom)

The approach was applied to EEG data originating from the CODYSEP dataset, designed to study the impact of sclerosis in inter-hemispherical transfer. The dataset consists in 31 ill subjects (hereafter termed SEP) and 20 controls (TEM); 17 channels EEG signals were collected at a 256 Hz sampling rate.

A subset of 14 ill patients and 16 controls was selected, namely those exhibiting sufficiently similar time-frequency contents, in particular in the 8-12 Hz range. For the sake of precision in the time domain, the hop size N_f was set to $N_f = 8$ (i.e. a time resolution of 125 ms), resulting in a moderate frequency resolution (each MDCT basis function having a bandwidth of approximately 4Hz).
This choice is certainly questionable, and the selection procedure should probably turned into a more sustematic one. Our choice was motivated by the desire of focusing on th same phenomenon (i.e. wiht the same time-frequency patterns) for all retained subjects.

3.2 Estimation results

A subset of sensors was also selected that correspond to relevant scalp locations for observing alpha waves. Corresponding MDCT transforms were computed, and the model parameters (for a two-state model: alpha and non alpha states) were estimated using the algorithms above. Examples of covariance matrices for the two states are displayed in Fig. 4 and 5. The frequency covariance matrices shown in Fig. 4 are almost diagonal, which indicates a decorrelation of the frequency bands. The main difference between the two matrices appears for the frequency band 8-12Hz (element (3,3)), which is significantly bigger in 'alpha' state. This state therefore succeds to capture alpha waves. The channel covariance matrices in Fig 5 are also significantly different for some channels: in particular, channels O1 and O2 (matrix elements (4,4) and (11,11)) are overactivated in 'alpha' state. Again this is coherent, since those sensors correspond to regions where alpha wave signals are most visible. This phenomenon is even more obvious on the graphical representation of topographies (Fig. 6),actually representing the diagonals of matrices of Fig 5 on which the relevant sensors O1 and O2 are clearly more "energetic" in 'alpha' state.

After completion of parameter estimation, a maximum likelihood estimate for the sequence of latent states is obtained via the Viterbi algorithm.

3.3 A study of alpha wave desynchronization

Let us start by quoting [3]: "*in multiple sclerosis, both axonal damage and demyelination occur. Therefore it can be expected that the connectivity between the different regions in the brains of MS patients will be impaired compared to healthy controls.*" Losses in

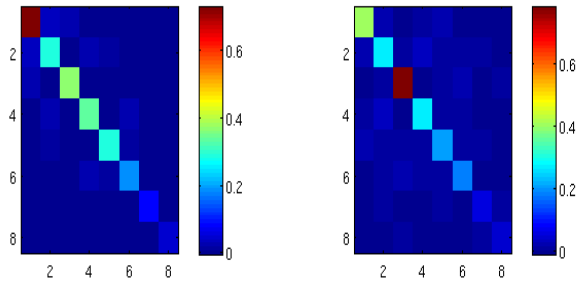

Figure 4: Frequency covariance matrices for state 'non-alpha' (left) and 'alpha' (right)

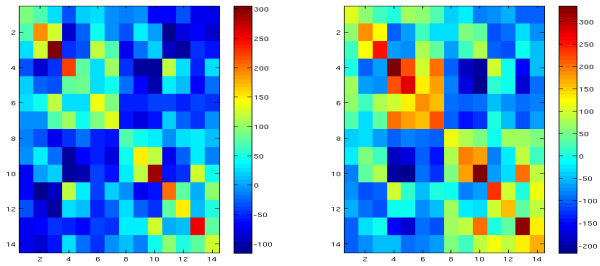

Figure 5: Channel covariance matrices for state 'non-alpha' (left) and 'alpha' (right)

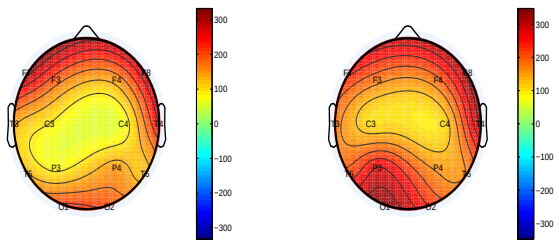

Figure 6: Representations of the channels variances for state 'non-alpha' (left) and 'alpha' (right)

synchronization between EEG signals in left and right hemispheres have been reported in the literature, where coherence is generally used to detect the phenomenon. The HHMM method described here is an alternative natural candidate for testing and evaluating it.

To test the desynchronization assumption, hidden states sequences were estimated for all subjects in the dataset, separately for left sensors and rignt sensors. The latter sequences take the form of sequences of zeros (non-alpha state) and ones (alpha state). For each subject, the Hamming distance between the left sensors $X_t^{(L)}$ and right sensors $X_t^{(R)}$ sequences was computed

$$D = \sum_t \left| X_t^{(L)} - X_t^{(R)} \right|$$

and the empirical distributions of the so-obtained distances were compared using a non parametric test. The figure 7 shows the boxplots of the two families (SEP and TEM) of distances, and seems to indicate a significant difference: the SEP data exhibit a larger left-right assymetry, in accordance with the above mentioned desynchronization hypothesis. To assess statistically the significance of the result, we performed a Mann-Whitney test. The corresponding

P-value was found to equal $P = 0.0384$, which confirms quantitatively the hypothesis of two distinct distributions.

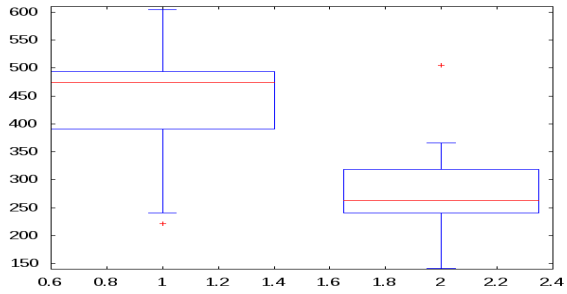

Figure 7: Boxplots of the Hamming distances between left and right hidden states for SEP (left) and TEM (right)

4. CONCLUSIONS AND PERSPECTIVES

We have described in this paper a new stochastic signal model, based on hidden Markov modelling of vectors of MDCT coefficients of multichannel signals. We have described corresponding estimation algorithms, and an adaptation to the case where covariance matrices take the form of Kronecker products of smaller matrices. Finally, we have tested the algorithms on a case study involving EEG signals.

While the algorithms described here are not new, the model itself and its adaptation to EEG type data is original. Our numerical results on simulations show that the estimators are good enough to handle data with realistic characteristics (length, number of channels,...).

In addition, the case study clearly shows the relevance of the model for real data processing, since it yields a biologically relevant result. Further work will concern the validation of this approach to different types of EEG detection and characterization problems. Applications to BCI (brain computer interfaces) problems are currently under study. Such applications will probably require online versions of the algorithms.

From a more fundamental point of view, it is to be noted that MDCT bases are known for having important limitations in terms of time-frequency resolution. It is therefore desirable to turn to better quality time-frequency expansions, such as Gabor frames. However, the loss of orthogonality in the waveform system makes the estimation more complex, as it will not be possible any more to use the expansion as an orthogonal transform and develop standard algorithms on the expansion coefficients. Dedicated algorithms have to be designed, which is work in progress.

Acknowledgements

This work was partly supported by CNRS, through the *NeuroInformatique* program's project **Codysep**, which allowed the recording of the data used in this work, and by the ANR project **CO-ADAPT** (ANR-09-DEFIS-002).
We wish to thank the anonymous reviewers for their very relevant comments, which helped us improving the paper.

REFERENCES

[1] F. Bijma and J. C. De Munck. A space-frequency analysis of MEG background processes. *Neuroimage*, 43:478–488, 2008.

[2] R. Carmona, W. Hwang, and B. Torrésani. *Practical Time-Frequency Analysis: continuous wavelet and Gabor transforms, with an implementation in S*, volume 9 of *Wavelet Analysis and its Applications*. Academic Press, San Diego, 1998.

[3] K. S. Cover, H. Vrenken, J. J. Geurts, B. W. van Oosten, B. Jelles, C. H. Polman, C. J. Stam, and B. W. van Dijk. Multiple sclerosis patients show a highly significant decrease in alpha band interhemispheric synchronization measured using MEG. *NeuroImage*, 29(3):783–788, February 2006.

[4] P. Flandrin. *Temps-Fréquence*. Traité des Nouvelles Technologies, série Traitement du Signal. Hermès, Paris, 1993.

[5] M. Fornasier and H. Rauhut. Recovery algorithms for vector valued data with joint sparsity constraints. *SIAM Journal on Numerical Analysis*, 46(2):577–613, 2008. to appear.

[6] M. Kowalski and B. Torrésani. Décompositions parcimonieuses et persistantes de signaux multicanaux. applications aux signaux MEEG. In F. d'Alché Buc, editor, *Proceedings of CAP'08, 10eme Conference d'Apprentissage Porquerolles*, pages 105–120, Porquerolles, France, May 2008. Cepadues Editions.

[7] S. Molla and B. Torrésani. An hybrid audio scheme using hidden Markov models of waveforms. *Applied and Computational Harmonic Analysis*, 18(2):137–166, 2005.

[8] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.

[9] P. Soendergaard. LTFAT, the linear time-frequency analysis toolbox, 2010.

[10] D. Studer, U. Hoffmann, and T. Koenig. From EEG dependency multichannel matching pursuit to sparse topographic EEG decomposition. *Journal of Neuroscience Methods*, 153(2):261–275, June 2006.